DTU Library

Impact of background noise and sentence complexity on cognitive processing demands

Wendt, Dorothea; Dau, Torsten; Hjortkjær, Jens

Published in:
Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling

Publication date:
2015

Document Version
Publisher's PDF, also known as Version of record

Citation (APA):
Wendt, D., Dau, T., & Hjortkjær, J. (2015). Impact of background noise and sentence complexity on cognitive processing demands. In S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, & T. Andersen (Eds.), *Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling*

Impact of background noise and sentence complexity on cognitive processing demands

DOROTHEA WENDT[1,2,*], TORSTEN DAU[1], AND JENS HJORTKJÆR[1,3]

[1] *Hearing Systems Group, Department of Electrical Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark*

[2] *Eriksholm Research Centre, Snekkersten, Denmark*

[3] *Danish Research Centre for Magnetic Resonance, Copenhagen University Hospital Hvidovre, Denmark*

Speech comprehension in adverse listening conditions requires cognitive processing demands. Processing demands can increase with acoustically degraded speech but also depend on linguistic aspects of the speech signal, such as syntactic complexity. In the present study, pupil dilations were recorded in 19 normal-hearing participants while processing sentences that were either syntactically simple or complex and presented in either high- or low-level background noise. Furthermore, the participants were asked to rate the subjectively perceived difficulty of sentence comprehension. The results showed that increasing noise levels had a greater impact on the perceived difficulty than sentence complexity. In contrast, the processing of complex sentences resulted in greater and more prolonged pupil dilations. The results suggest that while pupil dilations may correlate with cognitive processing demands, acoustic noise has a greater impact on the subjective perception of difficulty.

INTRODUCTION

Everyday listening situations usually take place at high signal-to-noise ratios (SNR), ranging from +5 to +15 dB (Smeds *et al.*, 2015). Nevertheless, in some situations, listeners may experience considerable difficulties with listening to speech even though intelligibility is at 100%. The processing demands might be high in such situations and comprehension may be experienced as effortful. The listening difficulties may arise from the acoustic disturbance of the speech source due to the background noise, or caused by a hearing impairment, but may further result from purely endogenous factors, such as the complexity of the speech signal that is being processed. While both acoustic and cognitive factors may challenge the processing load, it is still unknown whether they interact in the experience of listening effort. Different measures have been used in order to investigate effortful listening, ranging from subjective measures, such as subjectively rated effort, to more objective or physiological measures, such as task-evoked pupil dilation as an indicator of increased cognitive processing demands (McGarrigle *et al.*, 2014).

*Corresponding author: wendt@elektro.dtu.dk

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

Several studies have demonstrated a correlation between pupil dilations and task demands (Beatty, 1982; Kahneman and Beatty, 1966). Thus, pupillometry has been increasingly used to examine processing load (sometimes termed ‘listening effort’) during speech recognition in difficult listening environments (Zekveld *et al.*, 2010; 2011). For example, Zekveld *et al.* (2010) examined the pupil response of listeners with normal hearing who listened to sentences presented in noise at several signal-to-noise ratios. They reported that mean pupil dilation and peak pupil dilation increased with increasing noise level indicating higher processing demands. Besides background noise, a few studies indicated that linguistic aspects of the speech signal, such as syntactic complexity, affect speech processing. More complex sentence structures can lead to a decrease in speech intelligibility (Uslar *et al.*, 2013) or an increase in processing duration (Wendt *et al.*, 2014; 2015). Piquado *et al.* (2010) used the pupillary response in younger and older adults to test cognitive processing demands due to syntactically complex sentences and sentence length. They found that the pupil response correlated with the length of the sentences, especially for elderly people. These studies indicated that processing demands are substantially higher when processing linguistically complex sentences in noise than when processing sentences with a simple linguistic structure.

The relationship between these subjective and objective measures of processing demands is still not well established (e.g., McGarrigle *et al.*, 2014). Although both measures have been employed, it seems that perceived effort and pupil dilation are not necessarily correlated (see, e.g., Zekveld *et al.*, 2011). In the present study, we examined the effects of syntactic complexity and noise level on processing demand using both a subjectively rated difficulty measure and pupil dilation. The rationale behind combining different measures in an audio-visual picture-matching task was to better understand the relationship between subjectively perceived difficulty and a physiological measure of effort.

MATERIAL AND METHODS

Participants

Eleven female and eight male participants with normal hearing carried out the experiment, with an average age of 23 years (ranging from 19 to 36 years). The participants had pure-tone hearing thresholds of 15 dB hearing level (HL) or better at the standard audiometric frequencies in the range from 125 to 8000 Hz. All participants had normal or corrected-to-normal vision.

Material

Speech material. Two different sentence types were recorded by translating 39 items from the German OLACS corpus (see Uslar *et al.*, 2013). All sentences contained a transitive full verb, an auxiliary verb (*vil* – ‘will’), a subject noun phrase (SNP) and an object noun phrase (ONP). Two different types of sentence structures were realized by varying the word order to either subject-verb-object structure (SVO) or to object-verb-subject structure (OVS). For each sentence structure, two different propositions were realized (*bjørn* – 'bear' as agent vs. *robot* – 'robot' as agent; see SVO I and OVS II in Table 1). The word order (the position of the main verb, e.g., *vække* – 'wake up') was the only cue to understanding *who* (the agent, i.e., the entity that carries out the action) *did what to whom* (the patient, i.e., the entity that is affected by the action). Both sentence structures (SVO and OVS) were locally ambiguous with respect to their meaning as well as to the grammatical role of the involved entities (e.g., *bjørn* and *robot* in Table 1) until after the auxiliary verb (*vil*).

Sentence type	**Example**							
	Word 1	Word 2	Word 3	Word 4	Word 5	Word 6	Word 7	Word8
SVO I	Den	flinke	bjørn	vil	$\text{vække}_{\text{PTD}}$	den	rare	robot.
	The agile bear will wake up the nice robot.							
SVO II	Den	rare	robot	vil	$\text{vække}_{\text{PTD}}$	den	flinke	bjørn.
	The nice robot will wake up the agile bear.							
OVS I	Den	flinke	bjørn	vil	den_{PTD}	rare	robot	vække.
	The agile bear, the nice robot will wake up.							
OVS II	Den	rare	robot	vil	den_{PTD}	flinke	bjørn	vække.
	The nice robot, agile bear will wake up.							

Table 1: Examples of the two different sentence structures that were used in the current study, i.e. subject-verb-object structure (SVO) and object-verb-subject structure (OVS). PTD indicates the point of target disambiguation.

In both structures, the disambiguating word, which is the word that allows a thematic role assignment of the agent and the patient – *who* (agent) *did what to whom* (patient) – is the auxiliary verb (see word 5 in Table 1). For instance, the position of the verb *vække* of the SVO structure (see SVO I and II in Table 1) disambiguates the sentence in a way that enables the participants to relate the spoken sentence to the target picture. For the OVS structure, the lack of a main verb in front of the article *den* (Table 1) informs the participants about the object role of the first noun in the sentence. Therefore, the onset of word 5 was defined as the "point of target disambiguation" (PTD). The SVO structure is considered syntactically simple and easy to process. Written and spoken OVS clauses in Danish, however, are typically more difficult to process (see Kristensen, 2013).

Visual stimuli. For each spoken sentence, a single picture was shown, which was either a target picture or competitor picture. The target picture illustrated the situation described by the spoken sentence (see right picture in Fig. 1). In the competitor picture, the roles of the agent and the object were interchanged (left picture in Fig. 1).

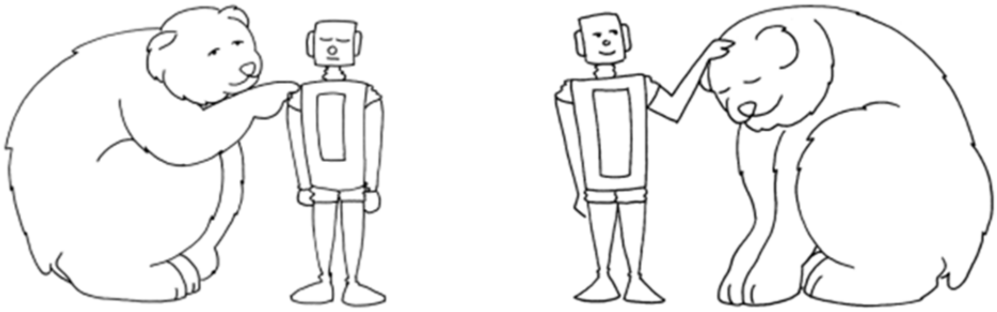

Fig. 1: Example of a target picture (right) and a competitor picture (left) for the Danish versions of the sentence "*The nice robot will wake up the agil bear*" or "*The agile bear, the nice robot will wake up.*"

PROCEDURE AND DATA ANALYSIS

All participants performed an audio-visual matching paradigm (see Fig. 2). First, a picture was shown on the screen for 2000 ms. The participants then heard the sentence (e.g., *The nice robot will wake up the agile bear*) while presented with a fixation cross. The sentences began 3000 ms after the picture offset and were presented in background noise. The background noise started 3000 ms before the sentence onset and ended 3000 ms after the sentence. After the noise offset, the participants' task was to decide whether the sentence matched the picture or not. 12 filler trials were included that did not contain a target or competitor picture, but showed an unrelated picture depicting different characters or different actions. After the comprehension question, the participants were asked to rate the perceived difficulty on a rating scale, i.e., how difficult they perceived the sentence comprehension to be. First, the participants performed one training block, which contained 10 trials. Afterwards, each participant listened to 159 sentences, divided into two blocks. The sentences were presented either at a low-noise level (+12 dB SNR) or at a high-noise level (−6 dB SNR). The noise masker was a stationary speech-shaped noise with the long-term frequency spectrum matching that of the speech. Changes in pupil size were measured for each participant from the onset of the noise until the comprehension task. An eye-tracker system (EyeLink 1000 desktop system, SR Research Ltd.) was used with a sampling rate of 1000 Hz to record pupil dilations.

Pupil data analysis. The pupil data were analysed using a similar procedure as described in Piquado *et al.* (2010) and Zekveld *et al.* (2010; 2011). First, the pupil data were cleaned for eye-blinks by classifying samples as an eye-blink for which the pupil value was below 3 standard deviations of the mean. Eye-blinks were removed and linearly interpolated, starting ten samples before and ending twenty samples after a blink. Trials for which more than 20% of the data required an interpolation were removed from further data analysis. The data of the de-blinked

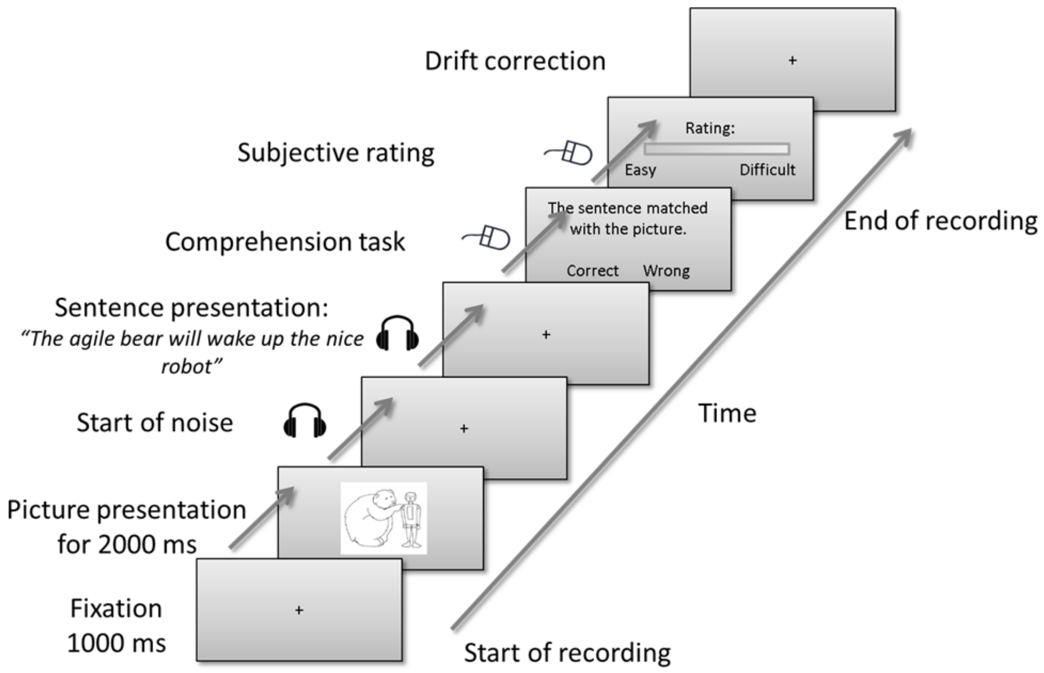

Fig. 2: Audio-visual picture-matching paradigm used for recording pupil dilation and subjectively perceived difficulty. Participants were presented a picture, followed by a spoken sentence. Their task was to decide whether the sentence either did or did not match with the picture. The comprehension question was followed by a subjective rating of the difficulty of the task.

trials were smoothed by a four-point moving average filter and then averaged for each condition and participant. In order to control for individual differences in pupil range, the minimum pupil value of the entire trial time series (from trial onset to the comprehension task) was subtracted from each trial data point. Afterwards, the pupil data were divided by the range of the pupil size within the entire trial. This method was applied to ensure consistent scaling of the range of the pupil value between 0 and 1 within each trial. Finally, the pupil data were normalized by subtracting a baseline-value which was defined as the averaged pupil value across 1.5 seconds before sentence presentation (when listening to noise alone). The maximum pupil dilation and time-averaged pupil dilation was calculated for the time interval between the sentence onset until the comprehension question.

RESULTS

A two-way repeated measures analysis of variance (ANOVA) was applied on the pupil data (on both mean pupil dilation and maximum pupil dilation) and the rated effort separately using SPSS 20, with *complexity* and *noise* as within-subjects factors. Significant effects were followed up with pairwise comparisons using post-hoc tests (applying a Bonferroni correction).

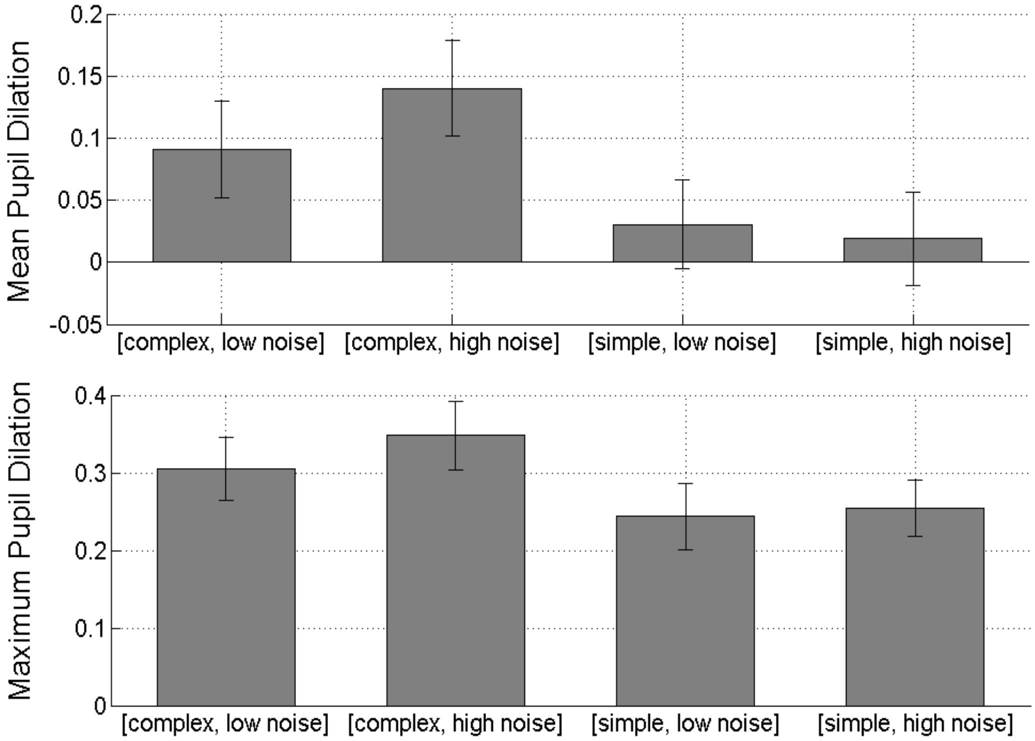

Fig. 3: Normalized pupil dilation averaged across all participants for four different conditions. The mean pupil dilation was calculated from the onset of sentence presentation until about 7500 ms after the sentence onset (3000 ms after sentence offset). Error bars indicate standard deviations.

Mean pupil dilation. A significant main effect was observed for the factor *complexity*. Post-hoc tests revealed significant differences in mean pupil dilation when processing syntactically simple and complex sentences [$F(1,18) = 22.0$, $p < 0.001$]. However, no interaction between noise and complexity was found.

Maximum pupil dilation. The ANOVA revealed a main effect of *complexity* [$F(1,18) = 13.0$, $p = 0.002$] indicating higher pupil dilation when processing syntactically complex sentences.

Difficulty rating. A significant main effect was found for the factor *noise*. Post-hoc tests revealed significant differences in pupil response in low and in high noise levels [$F(1,18) = 16.0$, $p < 0.001$]. This indicates that participants perceived processing sentences as more effortful within higher noise levels. The effect of sentence complexity, however, was rather small.

DISCUSSION

This study investigated subjective and physiological effects of linguistic complexity and background noise on sentence processing using an audio-visual picture-matching paradigm. Sentence processing demands were tested at two different levels

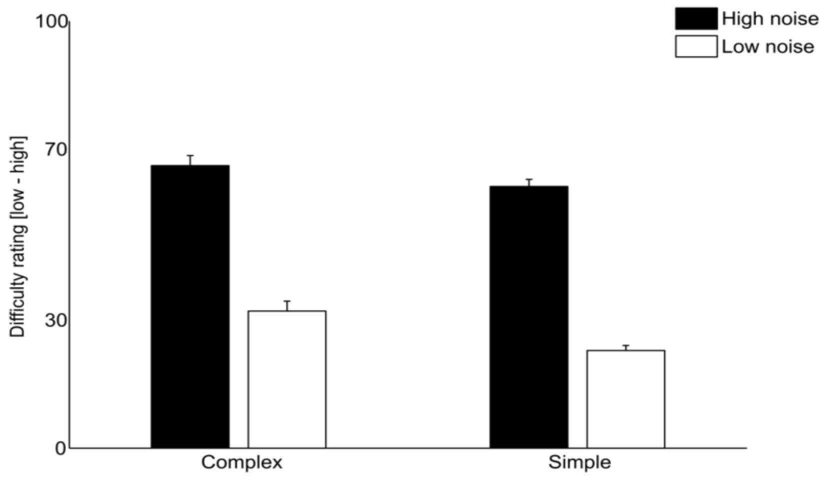

Fig. 4: Rated difficulty after the audio visual task averaged across all participants when OVS sentences (complex) or SVO (Simple) sentences were presented in low noise level (white) and high noise levels (black). Error bars show standard errors.

of background noise, whereby speech intelligibility was always relatively high. The results suggest that noise level and syntactic complexity relate in different ways to subjectively perceived effort and physiological markers of speech processing. The syntactic complexity was found to increase pupil dilation while processing sentences in background noise. However, the effect of the background noise level on the mean pupil dilation was rather small. An increase of the processing demand due to higher a noise level was only reflected in the subjective ratings. In other words, the poorer acoustical speech signal (due to the presence of background noise) led to a higher perceived demand on sentence processing. However, the interaction between background noise level and sentence structure was rather small. No combined effect of complexity and noise on either task-evoked pupil response or on perceived difficulties was found.

Our data indicate that both noise-induced and speech-induced processing demands can be found in listening situations that reflect everyday communication situations when speech intelligibility is still high. Moreover, our results demonstrate that the subjectively perceived effort is not directly reflected by the pupil dilation. The subjectively rated effort was found to be sensitive to changes in the noise level and, therefore, may reflect sensory processing difficulties that occur at early stages of speech processing (bottom-up processes). In contrast, the pupil response, which is often used as an indicator of the listening effort (Zekveld *et al.*, 2010), seems to be more sensitive to syntactic complexity and, thus, may reflect demands associated with cognitive processes that are required for sentence comprehension (top-down processes).

ACKNOWLEDGMENTS

This project was funded by the Oticon Centre of Excellence for Hearing and Speech Sciences (CHeSS). Thanks go to Line Burholt Kristensen for helping with the development of the sentence material. We thank the AULIN project for using their picture material.

REFERENCES

Beatty, J. (**1982**). "Task-evoked pupillary responses, processing load, and the structure of processing resources," Psychol. Bull., **91**, 276-292.

Kahneman, D. and Beatty, J. (**1966**). "Pupil diameter and load on memory." Science, **154**, 1583-1585.

Kristensen, L.B. (**2013**) *Context, You Need: Experimental Approaches to Information Structure Processing.* PhD thesis, University of Copenhagen.

McGarrigle, R., Munro, K.J., Dawes, P., Stewart, A.J., Moore, D.R., Barry, J.G., and Amitay, S. (**2014**). "Listening effort and fatigue: What exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group 'white paper'," Int. J. Audiol., **53**, 433-445.

Piquado, T., Isaacowitz, D., and Wingfield, A. (**2010**). "Pupillometry as a measure of cognitive effort in younger and older adults," Psychophysiology, **47**, 560-569.

Smeds, K., Wolters, F., and Rung, M. (**2015**). "Estimation of Signal-to-Noise Ratios in Realistic Sound Scenarios." J. Am. Acad. Audiol., **26**, 183-196.

Uslar, V.N., Carroll, R., Hanke, M., Hamann, C., Ruigendijk, E., Brand, T., and Kollmeier, B. (**2013**). "Development and evaluation of a linguistically and audiologically controlled sentence intelligibility test," J. Acoust. Soc. Am., **124**, 3039-3056.

Wendt, D., Brand, T., and Kollmeier, B. (**2014**). "An eye-tracking paradigm for analyzing the processing time of sentences with different linguistic complexities," PLoS ONE, **9**, e100186.

Wendt, D., Kollmeier, B., and Brand, T. (**2015**). "How hearing impairment affects sentence comprehension: Using eye fixations to investigate the duration of speech processing," Trends Hear., **19**, 1-18.

Zekveld, A.A., Kramer, S.E., and Festen, J.M. (**2010**). "Pupil response as an indication of effortful listening: The influence of sentence intelligibility," Ear Hearing, **31**, 480-490.

Zekveld, A.A., Kramer, S.E., and Festen, J.M. (**2011**). "Cognitive load during speech perception in noise: the influence of age, hearing loss, and cognition on the pupil response," Ear Hearing, **32**, 498-510.